

PEER RANKING: Un nuevo enfoque para la evaluación formativa

M Rosario Perello-Marin^a, Jose Pedro Garcia-Sabater^b, Juan A. Marin-García^c y Joan Morant-Llorca^d

^aDepartamento de Organización de Empresas, Universitat Politècnica de València (Spain), rperell@upvnet.upv.es, ^bGrupo ROGLE. Universitat Politècnica de València (Spain), jpgarcia@omp.upv.es, ^cGrupo ROGLE. Universitat Politècnica de València (Spain), jamarin@omp.upv.es, ^dGrupo ROGLE. Universitat Politècnica de València (Spain), jmorantllorca@gmail.com

Abstract

This paper introduces peer ranking as a type of a formative assessment methodology. By using this methodology, students compare written reports from their peer in pairs. We explain the advantages of using this methodology not only to enhance student's learning but also as a way to assess student's transversal competence of critical thinking.

Keywords: Peer ranking, formative assessment, critical thinking, transversal competences.

Resumen

Este artículo introduce peer ranking como un tipo de metodología para la evaluación formativa. Mediante la aplicación de esta metodología, los alumnos deben comparar por pares, trabajos escritos elaborados por otros compañeros. En el presente trabajo se exponen las ventajas de la adopción de esta metodología no sólo para mejorar el aprendizaje de los estudiantes, sino también como un modo de evaluar la competencia transversal de pensamiento crítico.

Palabras clave: peer ranking, evaluación formativa, pensamiento crítico, competencias transversales.

1. Introducción

Uno de los principales objetivos del marco del Espacio Europeo de Educación Superior es el acompañamiento de los estudiantes en su proceso de aprendizaje a medida que ellos mismos van viendo su progreso en la construcción de sus conocimientos (Reinalda & Publishers, 2005). Esto es lo que se denomina evaluación formativa, es decir, un método de evaluación que proporciona a los alumnos retroalimentación durante todo el curso ayudándoles así a mejorar su aprendizaje, independientemente de que se otorguen calificaciones a sus trabajos o no (Higgins & Grant, 2010; Nicol & Macfarlane-Dick, 2006). De este modo los estudiantes refuerzan su propio aprendizaje como fruto de los resultados de sus evaluaciones.

Sin embargo, la realidad a la que nos enfrentamos en las aulas hoy en día es que los sistemas de evaluación no siempre son evaluaciones formativas. Por lo tanto, no siempre garantizan un mayor nivel de aprendizaje de los alumnos. Aunque en la mayoría de los casos, hay un

aumento en el número de actos de evaluación que antes de la reforma educativa actual, este hecho no siempre conlleva asociado una mayor motivación, mayor participación o un mejor nivel en los resultados de aprendizaje obtenidos por los estudiantes (Vickerman, 2009).

El análisis de distintos sistemas de evaluación existentes, muestra que la evaluación por pares aumenta la motivación de los estudiantes, ya que sienten que participan activamente en todo el proceso, no sólo en el aprendizaje, sino también en la evaluación. Por otra parte, los estudiantes también aprenden a ser más críticos, permitiéndoles comparar sus evaluaciones con las realizadas por otros compañeros o por los propios profesores. Diferentes autores confirman que la evaluación por pares enfatiza habilidades, fomenta la participación, se centra en el aprendizaje, promueve la excelencia, ofrece una mayor retroalimentación, fomenta la asistencia a clase, y entrena a los estudiantes en responsabilidad y pensamiento crítico (Ohland et al., 2012; Reiter, Eva, Hatala, & Norman, 2002; Sluijsmans, Dochy, & Moerkerke, 1998; Tu & Lu, 2005).

De todos los métodos de evaluación por pares existentes, a lo largo de este trabajo nos centramos en 'Peer ranking' como una metodología de evaluación a través de la comparaciones por pares, de parejas de elementos (sean trabajos, alumnos, ejercicios o cualquier otro objeto). Es un tipo específico de evaluación a través de la cual los trabajos se ordenan de mejor a peor (Reiter et al., 2002; Tu & Lu, 2005), aunque no es evidente la posición respecto a una norma (nota) que le correspondería a cada elemento, solo su posición relativa frente a otros elementos (Dolan et al., 2007; Marin-Garcia et al., 2014).

En el trabajo mostramos la aplicación de la metodología 'PEER RANKING' para apoyar la evaluación por pares como modo de mejorar el aprendizaje de los alumnos. que además facilita la objetividad de los resultados. Por otra parte, la metodología propuesta permite evaluar la competencia transversal de pensamiento crítico de una manera sencilla.

2. Objetivos

El objetivo del presente trabajo es describir una metodología, transferida de otras áreas de conocimiento y hasta ahora poco empleada en el mundo académico, 'PEER RANKING'.

En este trabajo no se pretende describir cómo se califican los trabajos de un determinado acto de evaluación. Por el contrario, dado un orden, previamente establecido por el profesor, se propone la adopción de esta metodología con el fin de mejorar sus resultados de aprendizaje, incrementando la participación y la motivación, al tiempo que podemos evaluar la competencia transversal de pensamiento crítico.

Se presenta un algoritmo que selecciona los trabajos a comparar por parte de cada alumno, ofreciendo diferentes niveles de dificultad en las comparaciones. Dicha herramienta muestra al alumno el resultado correcto al final del proceso y genera una calificación del alumno evaluador en función de la bondad de sus respuestas que podría emplearse como evaluación de su competencia transversal de pensamiento crítico.

3. Desarrollo de la innovación

La evaluación del alumno puede llevarse a cabo usando distintos métodos de evaluación en función del objetivo perseguido, y de lo que se quiera evaluar: conocimientos, competencias, o incluso en función de la materia a evaluar. En este trabajo se presenta una metodología que además de desarrollar el aprendizaje mediante la evaluación de los trabajos de los compañeros, facilita la evaluación de la competencia de pensamiento crítico a través de la evaluación de trabajos escritos. Esta metodología puede emplearse también para la evaluación de otros tipos de trabajos desarrollados por los alumnos, pero hemos elegido trabajos escritos como prueba piloto (Lai & Hwang, 2015).

El sesgo involuntario y la incertidumbre constituyen una parte indispensable de muchos procesos de toma de decisiones. En la mayor parte de las ocasiones están relacionados con la especificación de un problema de decisión, el entorno en el que la decisión tiene que ser hecha, y el carácter del sistema de valores y preferencias de un tomador de decisiones. La complejidad de este problema ha llevado al desarrollo de un marco para el análisis de robustez, es decir, una base teórica y una diversidad de métodos múltiples de apoyo dedicados a criterios de decisión que tomen en cuenta las incertidumbres internas y externas observadas en las situaciones reales de decisión (Kadziński, Słowiński, & Greco, 2015). Existen diferentes enfoques para el estudio de este tipo de problemas minimizando el sesgo y optimizando los resultados (Adler, Friedman, & Sinuany-Stern, 2002). En nuestro trabajo, construimos a partir del trabajo previo realizado por Kadziński et al (Kadziński et al., 2015). Ellos se centran en problemas multicriterio de ordenación con evaluaciones determinísticas del desempeño, y modelizan las decisiones del tomador de decisiones con modelos aditivos de valores multi-atributo definidos a través de declaraciones de preferencias por pares holísticas (i.e. la alternativa a es (aunque sea débilmente) preferida sobre la alternativa b).

‘PEER RANKING’ consiste en evaluar trabajos mediante la comparación entre ellos dos a dos identificando cuál de ellos es mejor frente a un criterio de evaluación concreto. En este caso, se propone que sean los alumnos quienes evalúan con el objetivo de medir su capacidad de valorar crítica y objetivamente trabajos de compañeros. .

Establecer un ranking entre los resultados de los alumnos no es tarea fácil, especialmente cuando el número de resultados a analizar es elevado y las diferencias en cuestión de calidad entre ellos no son tan grandes. De hecho, en lugar de elegir una alternativa de entre un conjunto de opciones, u ordenar todas las alternativas en base a su deseabilidad, es mucho más simple empezar comparando alternativas de 2 en 2 (Hullermeier & Furnkranz, 2004). Es por ello que utilizar el método ‘PEER RANKING’, ayuda sensiblemente a simplificar el problema de toma de decisiones y al mismo tiempo garantiza la objetividad en los resultados.

Aplicando estos principios para el desarrollo de esta innovación docente, se presenta una herramienta informática que, partiendo de un conjunto de trabajos previamente evaluados por los profesores, les ofrece a cada alumno diferentes parejas de trabajos para que los clasifiquen comparándolos entre sí, teniendo en cuenta los mismos factores o criterios utilizados previamente por los profesores.

Dichos trabajos corresponden a un determinado acto de evaluación de una asignatura concreta. Previamente, el profesor evalúa los trabajos empleando una rúbrica determinada en función de la asignatura y el acto de evaluación particular del que se trate, otorgándole una calificación numérica a cada uno de los trabajos. El software traduce dicha calificación numérica en una de las categorías $Z_{i,k}$ mostradas en la Tabla 1.

Tabla 1. Categorías de los trabajos

Categoría $Z_{i,k}$	Descripción
5	Muy bueno
4	Bueno
3	Reguar
2	Mejorable
1	Muy Mejorable

Una vez asignado cada trabajo a una de las categorías, tras la evaluación realizada por el profesor mediante su esquema de rúbricas, la aplicación propondrá a los alumnos que comparen trabajos de sus compañeros.

El algoritmo selecciona, para cada alumno, parejas de trabajos, i, j , correspondientes a categorías diferentes (fruto de la la evaluación del profesor) para los diferentes criterios, a fin de que el alumno los compare en base a una serie de criterios.

Los criterios mostrados a los alumnos corresponden a los mismos criterios de la rubrica empleada previamente por el profesor. Sin embargo al alumno sólo se le muestra la descripción que llevaría a obtener la máxima calificación en cada ítem de la rúbrica.

Así, en función de la descripción facilitada de un determinado criterio k , que seleccione si el trabajo i es mejor que el j (resultado 1), si son igual de buenos (resultado 0) o si el trabajo i es peor que el j (resultado -1). De este modo, se obtiene, para cada alumno a , diferentes conjuntos de comparaciones que se representa como

$${}^l X_{i,j}^a = \begin{cases} -1 \\ 0 \\ 1 \end{cases} \quad (1)$$

Donde a corresponde al alumno, l corresponde a la comparación que realiza, k al criterio, i y j el par de trabajos comparados.

El software analiza la bondad de dichas comparaciones en función de los resultados previos introducidos por el profesor (Z) para cada trabajo en cada característica, teniendo en cuenta la dificultad de las comparaciones solicitadas al alumno. La puntuación que recibe el alumno se obtiene sumando los valores que la tabla de doble entrada adjunta (tabla 2) genera para el par de trabajos evaluados frente a la valoración que para cada criterio realiza el alumno.

Tabla 2. Criterios de puntuación para cada comparación

	$l y a$ $k X_{i,j}$		
$Z_i - Z_j$	1	0	-1
4	1	-3	-4
3	2	-2	-3
2	3	-1	-2
1	4	1	-1
0	0	2	0
-1	-1	1	4
-2	-2	-1	3
-3	-3	-2	2
-4	-4	-3	1

De ese modo, el software va proponiendo comparaciones de trabajos previamente clasificados en categorías diferentes para cada uno de los criterios a fin de valorar la capacidad crítica del alumno. Así se le expone al alumno a comparaciones de diversas dificultades (entre trabajos de categorías próximas y trabajos de categorías más distantes), lo cual garantiza la objetividad de los resultados obtenidos.

Previo al trabajo de comparación, el alumno recibe instrucciones específicas sobre la tarea que ha de realizar. En ellas se especifica que el software le ofrecerá comparaciones de trabajos hasta que, o bien haya realizado 10 comparaciones, o bien haya alcanzado un total de 10 puntos positivos acumulados. Durante todo el proceso, el alumno conoce la puntuación que lleva acumulada en cada momento.

Como resultado del proceso se obtiene una calificación del alumno evaluador que podría emplearse como evaluación de su competencia transversal de pensamiento crítico.

4. Resultados

La metodología que se ha explicado en el apartado anterior, ha sido adoptada para la asignatura Logística de Aprovisionamiento y Distribución de la titulación Máster Universitario Logística Aprovisionamiento y Distribución de la Universitat Politècnica de Valencia durante el curso 2015-2016 a modo de prueba piloto. En particular, se ha aplicado a trabajos escritos de ‘Desarrollo de modelos de negocio’ como datos de entrada para el proceso de evaluación formativa de la competencia de pensamiento crítico de los alumnos.

Existen múltiples modos de medir la calidad de los trabajos escritos. Para este trabajo, se han seleccionado algunos de los criterios expuestos por Watts et Al (2012).

Así, proponemos medir la calidad de los trabajos escritos en base a los siguientes criterios: 1) Propone soluciones apropiadas a lo que demanda la tarea (profundidad); 2) transmite efectivamente las ideas; 3) usa los recursos ingeniosamente (originalidad). Aunque se podrían añadir otros criterios, hemos estimado que con estos 3 es suficiente a fin de comparar los trabajos escritos de esta asignatura de dos en dos por parte de los alumnos y evaluar así su capacidad de pensamiento crítico.

En la figura 1 se muestra una imagen de lo que el alumno visualiza en el momento de realizar un par de comparaciones.

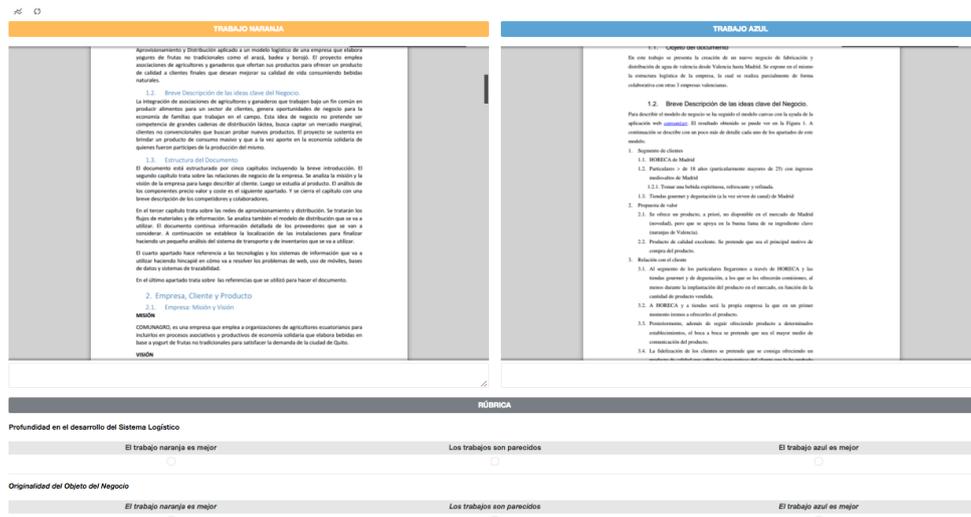


Fig 1. Ejemplo de un par de comparaciones

Al presentar esta herramienta a los alumnos, ellos se muestran receptivos ante esta innovación. Sin embargo, una de las barreras que se planteó fue el temor a que el resultado de su evaluación (la que realizaban los propios alumnos) pudiera afectar al resultado de evaluación (calificación) de sus compañeros. Este problema, se resolvió explicándoles que la evaluación del trabajo se había realizado con antelación.

Por otro lado, hubo alumnos que, en la mayoría de los casos, debido a la baja calidad de los trabajos que presentaron, no estuvieron de acuerdo con el hecho de que se hicieran públicos los trabajos, puesto que, pese a ser anónimos, ellos temían que fueran reconocibles. El mero hecho de que solicitaran que su trabajo quedara eliminado nos indica que la implementación de este tipo de actividades de evaluación puede ser un interesante factor de motivación y superación para los alumnos en futuras ediciones.

5. Conclusiones

Se ha mostrado 'PEER RANKING' como una metodología objetiva para la evaluación formativa. Con el apoyo de un algoritmo informático, el número de comparaciones que se le ofrecen al alumno para poder evaluar su capacidad de pensamiento crítico, están acotadas y garantizan que se le expone a pares de comparaciones de diversas dificultades velando así por la objetividad de los resultados de evaluación obtenidos.

Forzar al alumno a evaluar trabajos de sus pares de diferentes niveles se muestra, intuitivamente, como un modo de educar su pensamiento crítico. Además al haber una clasificación objetiva realizada previamente, el alumno tiene una presión adicional por hacer bien su trabajo evitando los riesgos de evaluar según afinidad o de evitar la comparación.

De este modo, se obtienen dos objetivos simultáneamente. Por un lado los alumnos mejoran sus resultados de aprendizaje comparando trabajos excelentes con otros no tan buenos. Para ello se les da la oportunidad de identificar las diferencias entre ellos así como las razones que los hacen peores o mejores. Y por otro lado, se obtiene una evaluación sencilla de la competencia transversal de pensamiento crítico.

De cara a investigaciones futuras, esta metodología puede extenderse a otro tipo de trabajos, no quedando restringida a trabajos escritos, y a otras áreas de conocimiento.

Sería interesante también encontrar el modo de que no hiciera falta una evaluación previa detallada para utilizar la evaluación de los alumnos como mecanismo para obtener la clasificación. Además este esquema podría facilitar la detección temprana de las desviaciones en la formación de su espíritu crítico.

Nuestra intención sería incluir este tipo de evaluaciones en las guías docentes de algunas asignaturas para el próximo curso.

6. Referencias

ADLER, N., FRIEDMAN, L., y SINUANY-STERN, Z. (2002). Review of ranking methods in the data envelopment analysis context. *European Journal of Operational Research*, 140(2), 249–265.

DOLAN, S. L., VALLE CABRERA, R., JACKSON, S. E., y SCHULER, R. S. (2007). *La gestión de los recursos humanos. Cómo atraer, retener y desarrollar con éxito el capital humano en tiempos de transformación*. Madrid: McGraw-Hill.

HIGGINS, M., y GRANT, F. (2010). Formative Assessment : Balancing Educational Effectiveness and Resource Efficiency. *Education*, 5(2), 4–24.

HULLERMEIER, E., y FURNKRANZ, J. (2004). Ranking by Pairwise Comparison: A Note on Risk Minimization. *Fuzzy Systems, 2004. Proceedings. 2004 IEEE International Conference on*, 1, 97–102 vol.1.

KADZIŃSKI, M., SŁOWIŃSKI, R., y GRECO, S. (2015). Multiple criteria ranking and choice with all compatible minimal cover sets of decision rules. *Knowledge-Based Systems*, 89, 569–583.

LAI, C.-L., y HWANG, G.-J. (2015). An interactive peer-assessment criteria development approach to improving students' art design performance using handheld devices. *Computers & Education*, 85, 149–159.

MARIN-GARCIA, J. A., ARAGONÉS BELGRAN, P., & MELÓN, G. (2014). Intra-rater and inter-rater consistency of pair wise comparison in evaluating the innovation competency for university students. *Working Papers on Operations Management*, 5(2), 24-46.

NICOL, D. J., y MACFARLANE- DICK, D. (2006). Formative assessment and self- regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218.

OHLAND, M. W., LOUGHRY, M. L., WOEHR, D. J., BULLARD, L. G., FELDER, R. M., FINELLI, C. J., y SCHMUCKER, D. G. (2012). The Comprehensive Assessment of Team Member Development of a Behaviorally Anchored Rating Scale for Self- and Peer Evaluation. *Academy of Management Learning and Education*, 11(4), 609–631.

REINALDA, B., y PUBLISHERS, B. B. (2005). The Bologna Process — Harmonizing Europe ' s Higher Education.

REITER, H. I., EVA, K. W., HATALA, R. M., y NORMAN, G. R. (2002). Self and peer assessment in tutorials: application of a relative-ranking model. *Academic Medicine : Journal of the Association of American Medical Colleges*, 77(11), 1134-9.

SLUIJSMANS, D., DOCHY, F., y MOERKERKE, G. (1998). Creating a learning environment by using self-, peer-and co-assessment. *Learning Environments Research*, 293-319.

TU, Y., y LU, M. (2005). Peer-and-Self Assessment to Reveal the Ranking of Each Individual's Contribution to a Group Project. *Journal of Information Systems Education*, 16(2), 197-205.

VICKERMAN, P. (2009). Student perspectives on formative peer assessment: an attempt to deepen learning? *Assessment & Evaluation in Higher Education*, 34(2), 221-230.

WATTS, FRANCES, MARIN-GARCIA, JA, GARCÍA CARBONELL, A Y AZNAR-MAS, LE (2012) "Validation of a rubric to assess innovation competence." *WPOM-Working Papers on Operations Management* 3(1), 61-70.